UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
10/600,475	06/20/2003	Chris J.C. Burges	MCS-018-03	6335

7590 06/22/2007

LYON & HARR, L.L.P
Suite 800
300 Esplanade Drive
Oxnard, CA 93036-1274

EXAMINER	
SIEDLER, DOROTHY S	
ART UNIT	PAPER NUMBER
2626	

MAIL DATE	DELIVERY MODE
06/22/2007	PAPER

Please find below and/or attached an Office communication concerning this application or proceeding.

The time period for reply, if any, is set in the attached communication.

Office Action Summary	Application No.	Applicant(s)	
	10/600,475	BURGES ET AL.	
	Examiner	Art Unit	
	Dorothy Sarah Siedler	2626	

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --

Period for Reply

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE 3 MONTH(S) OR THIRTY (30) DAYS, WHICHEVER IS LONGER, FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133). Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

Status

1)☒ Responsive to communication(s) filed on *18 May 2007*.

2a)☐ This action is **FINAL**. 2b)☒ This action is non-final.

3)☐ Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

Disposition of Claims

4)☒ Claim(s) *1-60* is/are pending in the application.

4a) Of the above claim(s) *30-60* is/are withdrawn from consideration.

5)☐ Claim(s) _____ is/are allowed.

6)☒ Claim(s) *1-29* is/are rejected.

7)☐ Claim(s) _____ is/are objected to.

8)☐ Claim(s) _____ are subject to restriction and/or election requirement.

Application Papers

9)☐ The specification is objected to by the Examiner.

10)☒ The drawing(s) filed on *20 June 2003* is/are: a)☒ accepted or b)☐ objected to by the Examiner.

Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).

Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).

11)☐ The oath or declaration is objected to by the Examiner. Note the attached Office Action or form PTO-152.

Priority under 35 U.S.C. § 119

12)☐ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).

a)☐ All b)☐ Some * c)☐ None of:

1.☐ Certified copies of the priority documents have been received.

2.☐ Certified copies of the priority documents have been received in Application No. _____.

3.☐ Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).

* See the attached detailed Office action for a list of the certified copies not received.

Attachment(s)

1) ☒ Notice of References Cited (PTO-892)
2) ☐ Notice of Draftsperson's Patent Drawing Review (PTO-948)
3) ☒ Information Disclosure Statement(s) (PTO/SB/08) Paper No(s)/Mail Date *11-3-03*.
4) ☐ Interview Summary (PTO-413) Paper No(s)/Mail Date. _____.
5) ☐ Notice of Informal Patent Application
6) ☐ Other: _____.

DETAILED ACTION

This is the initial office action in response to the application filled June 20, 2003. Claims 1-29 are pending and are considered below.

Applicant's election without traverse of group 1, including claims 1-29, in the reply filed on May 18, 2007 is acknowledged.

Claims 30-60 are withdrawn from further consideration pursuant to 37 CFR 1.142(b) as being drawn to a nonelected group 2, there being no allowable generic or linking claim. Election was made **without** traverse in the reply filed on May 18, 2007.

Claim Rejections - 35 USC § 112

The following is a quotation of the second paragraph of 35 U.S.C. 112:

> The specification shall conclude with one or more claims particularly pointing out and distinctly claiming the subject matter which the applicant regards as his invention.

Claim 18 is rejected under 35 U.S.C. 112, second paragraph, as being indefinite for failing to particularly point out and distinctly claim the subject matter which applicant regards as the invention. Claim 18 recites "omitting an output non-linearity, which was used during training..to generate a modified feature vector output", however this is unclear. The examiner acknowledges the description on the specification, however this does not clarify "omitting an output non-linearity". It is unclear as to what step or piece of information, used during training, is omitted. Therefore the examiner interprets claim 18 as the anchor model producing a likelihood score during the use phase. The likelihood

score is then normalized to produce a modified feature vector output. This interpretation is used throughout the remainder of this office action.

Claim Rejections - 35 USC § 103

The following is a quotation of 35 U.S.C. 103(a) which forms the basis for all obviousness rejections set forth in this Office action:

> (a) A patent may not be obtained though the invention is not identically disclosed or described as set forth in section 102 of this title, if the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which said subject matter pertains. Patentability shall not be negatived by the manner in which the invention was made.

Claims 1-4,6-10,12-14,17-19, and 20-28 are rejected under 35 U.S.C. 103(a) as being unpatentable over ***Sturim*** ("Speaker Indexing in Large Audio Databases Using Anchor Models" 2001) in view of ***Rudasi*** ("Text-Independent Talker Identification with Neural Networks" IEEE 1991).

As per claims 1 and 14, ***Sturim*** discloses a method for processing audio data, comprising: applying a plurality of anchor models to the audio data (Section 1. Introduction, *a target utterance is characterized using anchor models derived from a predetermined set of speakers*); mapping the output of the plurality of anchor models into frame tags, and producing the frame tags (section 2. Anchor Models, *speaker characterization vectors are mapped onto a speaker space. A determination of the speaker is made based on the location of the vector within speaker space*). ***Sturim*** does not disclose wherein the plurality of anchor models comprise a discriminatively trained classifier that is previously trained using a training technique. However, ***Sturim***

does disclose that anchor models, previously trained during a training phrase, can consist of any method of speech representation (section 1. Introduction and section 2. Anchor Models, first paragraph). In addition, ***Rudasi*** discloses a system for speaker identification and suggests that time-delay neural networks, a specific type of discriminately trained classifier, can be used (page 390, second paragraph). ***Sturim*** and ***Rudasi*** both disclose systems for the classification of speakers in audio data, and are therefore analogous art.

Therefore it would have been obvious to one of ordinary skill in the art at the time of the invention to use a discriminatively-trained classifier as an anchor model in ***Sturim***, since a time-delay neural network adds short term memory to the system, enabling the utilization of dynamic information which increases performance, as indicated in ***Rudasi*** (page 390, second paragraph).

As per claim 20, ***Sturim*** discloses a method for processing audio data containing a plurality of speakers, comprising: applying a plurality of anchor models to the audio data (Section 1. Introduction, *a target utterance is characterized using anchor models derived from a predetermined set of speakers*); mapping an output of the anchor models into frame tags (section 2. Anchor Models, *speaker characterization vectors are mapped onto a speaker space. A determination of the speaker is made based on the location of the vector within speaker space*); and a training set containing a set of training speakers, and wherein the plurality of speakers is not in the set of training

speakers (section 1. Introduction, *speakers of the target utterance are not members of the training* set). ***Sturim*** does not explicitly disclose constructing a list of start and stop times for each of the plurality of speakers based on the frame tags, and wherein the plurality of anchor models comprise a discriminatively-trained classifier previously trained using a training set. However, ***Sturim*** does disclose that anchor models, previously trained during a training phrase, can consist of any method of speech representation (section 1. Introduction and section 2. Anchor Models, first paragraph). In addition, ***Rudasi*** discloses a system for speaker identification and suggests that time-delay neural networks, a specific type of discriminately trained classifier, can be used (page 390, second paragraph). ***Sturim*** also discloses a system that can be used to retrieve messages from an archive (section 4. Speaker Indexing). In order to retrieve the messages, the system must know where they start and stop. Therefore the system must determine start and stop times for each speaker. ***Sturim*** and ***Rudasi*** both disclose systems for the classification of speakers in audio data, and are therefore analogous art.

Therefore it would have been obvious to one of ordinary skill in the art at the time of the invention to construct a list of start and stop times as well as use a discriminatively-trained classifier as an anchor model in ***Sturim***, since a time-delay neural network adds short term memory to the system, enabling the utilization of dynamic information which increases performance, as indicated in ***Rudasi*** (page 390, second paragraph). In addition, start and stop times can be used to reliably retrieve saved messages corresponding to a speaker for playback at a later time.

As per claims 2 and 17, ***Sturim*** in view of ***Rudasi*** disclose the method as set forth in claims 1 and 14, however neither explicitly disclose wherein the discriminatively-trained classifier comprises a convolutional neural network classifier. However, ***Rudasi*** discloses a system for speaker identification and suggests that time-delay neural networks can be used (page 390, second paragraph). In addition, by applicant's own admission, a time-delay neural network is a specific type of a convolutional classifier (Specification page 28, last paragraph).

Therefore it would have been obvious to one of ordinary skill in the art at the time of the invention to use a convolutional neural network, specifically a time-delay neural network, in ***Sturim***, since a time-delay neural network adds short term memory to the system, enabling the utilization of dynamic information which increases performance, as indicated in ***Rudasi*** (page 390, second paragraph).

As per claim 3, ***Sturim*** in view of ***Rudasi*** disclose the method as set forth in claim 2, and ***Sturim*** further comprising training the convolutional classifier on data separate from audio data available in a use phase (section 1. Introduction, *speakers of the target utterance are not members of the training* set).

As per claims 4 and 21, ***Sturim*** in view of ***Rudasi*** disclose the method as set forth in claims 2 and 20, and ***Rudasi*** further discloses wherein the convolutional neural network classifier is a time-delay neural network (TDNN) classifier (page 390, second paragraph).

Therefore it would have been obvious to one of ordinary skill in the art at the time of the invention to use time-delay neural network in ***Sturim***, since a time-delay neural network adds short term memory to the system, enabling the utilization of dynamic information which increases performance, as indicated in ***Rudasi*** (page 390, second paragraph).

As per claim 6, ***Sturim*** in view of ***Rudasi*** disclose the method as set forth in claim 1, and ***Rudasi*** further discloses pre-processing the audio data to generate input feature vectors for the discriminatively-trained classifier (page 390. Section 3.2 preprocessing of the acoustic data).

Therefore it would have been obvious to one of ordinary skill in the art at the time of the invention to pre-process the audio data to generate input feature vectors in ***Sturim***, since it would provide a reliable set of feature vectors, which can be easily applied to the classifier for further processing.

As per claims 7,8,18,19 and 22, ***Sturim*** in view of ***Rudasi*** disclose the method as set forth in claims 1,14 and 20 and ***Sturim*** further discloses omitting an output non-linearity, which was used during training, from the discriminatively-trained classifier, normalizing a feature vector output of the discriminatively-trained classifier, wherein the normalized feature vectors are vectors of unit length (section 2. Anchor Models, second paragraph, *each anchor model yields a likelihood score, where the combination of scores are used to form a N-dimensional characterization vector*, and the fifth paragraph

to the sixth paragraph, *a comparison is done between normalized data and non-normalized output data, therefore normalization must have been done*). In addition, Official notice is taken that it is old and well known to normalize a vector to a vector of unit length. Vectors are often normalized to a unit vector for simplicity of computation.

Therefore it would have been obvious to one of ordinary skill in the art at the time of the invention to normalize a vector output of the classifier to a unit vector and omit a non-linearity from the discriminatively-trained classifier in ***Sturim***, since it would produce a simplified feature vector, enabling simplified processing thus reserving computational resources.

As per claim 9, ***Sturim*** in view of ***Rudasi*** disclose the method as set forth in claim 1, and ***Rudasi*** further discloses accepting a plurality of input feature vectors corresponding to audio features contained in the audio data, and applying the discriminatively-trained classifier to the plurality of input feature vectors to produce a plurality of anchor model outputs (page 390, section 3.2 Preprocessing of the acoustic data).

Therefore it would have been obvious to one of ordinary skill in the art at the time of the invention to accept a plurality of input feature vectors corresponding to audio features contained in the audio data, and apply the discriminatively-trained classifier to the plurality of input feature vectors to produce a plurality of anchor model outputs in

Sturim, since it would provide a reliable set of feature vectors, which can be easily applied to the classifier for further processing.

As per claim 10, ***Sturim*** in view of ***Rudasi*** disclose the method as set forth in claim 1, and ***Sturim*** further discloses wherein the mapping comprises: clustering anchor model outputs from the discriminatively-trained classifier into separate clusters using a clustering technique, and associating a frame tag to each separate cluster (section 2. Anchor Models, *speaker characterization vectors are mapped onto a speaker space. A determination of the speaker is made based on the location of the vector within speaker space*).

As per claim 12, ***Sturim*** in view of ***Rudasi*** disclose the method as set forth in claim 1, and ***Sturim*** further discloses: training the discriminatively-trained classifier using a speaker training set containing a plurality of known speakers (section 1. Introduction, *anchor models are derived from a set of predetermined speakers*). ***Sturim*** does not explicitly disclose pre-processing the speaker training set and the audio data in the same manner to provide a consistent input to the discriminatively trained classifier. However, ***Rudasi*** discloses pre-processing of audio data (page 390, section 3.2 Preprocessing of the acoustic data).

Therefore it would have been obvious to one of ordinary skill in the art at the time of the invention to pre-process the speaker training set and the audio data in the same

manner in ***Sturim***, since it would provide reliable data input to the classifier, which would provide a reliable and useful result.

As per claims 13 and 23, neither ***Sturim*** in view of ***Rudasi*** explicitly disclose computer-readable medium having computer-executable instructions for performing the method recited in claims 1 and 20. However, the method of ***Sturim*** requires considerable computation and processing, and modern computer systems can perform the same computations considerably faster, and with higher accuracy, than any human would.

Therefore it would have been obvious to perform the method of claim 1 on a computer-readable medium in ***Sturim***, since a computer would enable faster processing, saving time and assuring accuracy.

As per claim 24, ***Sturim*** does not explicitly disclose a computer-readable medium having computer-executable instructions for processing audio data, comprising: training a discriminatively-trained classifier in a discriminative manner during a training phase to generate parameters that can be used at a later time by the discriminatively-trained classifier, applying the discriminatively-trained classifier that uses the parameters to the audio data to generate anchor model outputs; and clustering the anchor model outputs into frame tags of speakers that are contained in the audio data. However, ***Sturim*** does disclose training anchor models to be used to produce anchor

models outputs, and clustering anchor model outputs into frame tags of speakers (Section 1. Introduction, *a target utterance is characterized using anchor models derived from a predetermined set of speakers* and section 2. Anchor Models, *speaker characterization vectors are mapped onto a speaker space. A determination of the speaker is made based on the location of the vector within speaker space*). ***Sturim*** also discloses that anchor models, previously trained during a training phrase, can consist of any method of speech representation (section 1. Introduction and section 2. Anchor Models, first paragraph). In addition, ***Rudasi*** discloses a system for speaker identification and suggests that time-delay neural networks, a specific type of discriminately trained classifier, can be used (page 390, second paragraph)

Therefore it would have been obvious to one of ordinary skill in the art at the time of the invention to use a discriminatively-trained classifier as an anchor model in ***Sturim***, since a time-delay neural network adds short term memory to the system, enabling the utilization of dynamic information which increases performance, as indicated in ***Rudasi*** (page 390, second paragraph).

As per claim 25, ***Sturim*** in view of ***Rudasi*** disclose the computer-readable medium of claim 24, and ***Rudasi*** further discloses pre-processing a speaker training set during the training phase to produce a first set of input feature vectors for the discriminatively-trained classifier (page 390. Section 3.2 pre-processing of the acoustic data). However, neither ***Sturim*** nor ***Rudasi*** disclose pre-processing a speaker training set during a validation phase to produce a first set of input feature vectors for the

discriminatively-trained classifier. However, by applicants own admission (specification page 17, second paragraph) validation sets are old and well known.

Therefore it would have been obvious to one of ordinary skill in the art at the time of the invention to pre-process the audio data to generate input feature vectors in a training and validation phase in ***Sturim***, since it would provide a reliable set of feature vectors, which can be easily applied to the classifier for further processing.

As per claim 26, ***Sturim*** in view of ***Rudasi*** disclose the computer-readable medium of claim 25, further comprising pre-processing the audio data during the use phase to produce a second set of input feature vectors for the discriminatively-trained classifier, the pre-processing of the audio data being preformed in the same manner as the pre-processing of the speaker training set. ***Sturim*** does not explicitly pre-processing the audio data during the use phase to produce a second set of input feature vectors for the discriminatively-trained classifier, the pre-processing of the audio data being preformed in the same manner as the pre-processing of the speaker training set. However, ***Rudasi*** discloses pre-processing of audio data (page 390, section 3.2 Preprocessing of the acoustic data).

Therefore it would have been obvious to one of ordinary skill in the art at the time of the invention to pre-process the speaker training set and the audio data in the same manner in ***Sturim***, since it would provide reliable data input to the classifier, which would provide a reliable and useful result.

As per claims 27 and 28, ***Sturim*** in view of ***Rudasi*** disclose the computer-readable medium of claim 24, further comprising normalizing the feature vector outputs to produce feature vectors having a unit length, and omitting a nonlinearity from the discriminatively-trained classifier during the use phase (section 2. Anchor Models, second paragraph, *each anchor model yields a likelihood score, where the combination of scores are used to form a N-dimensional characterization vector*, and the fifth paragraph to the sixth paragraph, *a comparison is done between normalized data and non-normalized output data, therefore normalization must have been done*). In addition, Official notice is taken that it is old and well known to normalize a vector to a vector of unit length. Vectors are often normalized to a unit vector for simplicity of computation.

Therefore it would have been obvious to one of ordinary skill in the art at the time of the invention to normalize a vector output of the classifier to a unit vector and omit a nonlinearity from the discriminatively-trained classifier during the use phase in ***Sturim***, since it would produce a simplified feature vector, enabling simplified processing thus reserving computational resources.

Claims 5,15 and 16 are rejected under 35 U.S.C. 103(a) as being unpatentable over ***Sturim*** in view of ***Rudasi*** as applied to claims 4 and 14 above, and further in view of ***Lavagetto*** ("Time-Delay Neural Network for Estimating Lip Movements from Speech Analysis: A useful Tool in Audio-Video Synchronization" IEEE 1997).

Sturim in view of ***Rudasi*** disclose the method as set forth in claims 4 and 14, however neither explicitly disclose further training the TDNN classifier using cross entropy, nor training the classifier using a mean-square error metric. However, by applicant's own admission training using cross entropy is well known (specification page 29). In addition, ***Lavagetto*** discloses that training a time-delay neural network can be done with either cross entropy of mean-square error (page 789-790).

Therefore it would have been obvious to one of ordinary skill in the art at the time of the invention to use cross entropy or mean-square error to train the TDNN in ***Sturim*** and ***Rudasi***, since cross entropy and mean-square error provide figures for validating estimates provided by each network independent from the network structure itself, as indicated in ***Lavagetto*** (page 789,section IV. Learning Criteria for TDNN Training).

Claims 11 and 29 are rejected under 35 U.S.C. 103(a) as being unpatentable over ***Sturim*** in view of ***Rudasi*** as applied to claims 10 and 25 above, and further in view of ***Liu*** (6,615,170).

Sturim in view of ***Rudasi*** disclose the method as set forth in claims 10 and 25, however neither disclose applying temporal sequential smoothing to the frame tag using temporal information associated with the clustered anchor model outputs. ***Liu*** discloses temporal smoothing tagged frames (column 5 line 55- column 5 line 20). ***Liu*** discloses tagging speech frames based on the output of specific model. Adjacent observations are then used to update the vale of a tag for each frame by weighting observations at different times.

Therefore it would have been obvious to one of ordinary skill in the art at the time of the invention to apply temporal sequential smoothing to the frame tags in ***Sturim*** and ***Rudasi***, since it enables the incorporation of adjacent frame tags for updating and validating a current frame tag, thus increasing tagging accuracy, as indicated in Liu (column 5 lines 64-65).

Conclusion

The prior art made of record and not relied upon is considered pertinent to applicant's disclosure. Please see the PTO-892 form.

Any inquiry concerning this communication or earlier communications from the examiner should be directed to Dorothy Sarah Siedler whose telephone number is 571-270-1067. The examiner can normally be reached on Mon-Thur 9:30am-5:30pm.

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, Richemond Dorvil can be reached on 571-272-7602. The fax phone number for the organization where this application or proceeding is assigned is 571-273-8300.

Information regarding the status of an application may be obtained from the Patent Application Information Retrieval (PAIR) system. Status information for published applications may be obtained from either Private PAIR or Public PAIR. Status information for unpublished applications is available through Private PAIR only. For more information about the PAIR system, see http://pair-direct.uspto.gov. Should you have questions on access to the Private PAIR system, contact the Electronic Business Center (EBC) at 866-217-9197 (toll-free). If you would like assistance from a USPTO Customer Service Representative or access to the automated information system, call 800-786-9199 (IN USA OR CANADA) or 571-272-1000.

DSS

TĀLIVALDIS IVARS ŠMITS
PRIMARY EXAMINER